# 21. Exam 2 Review

Problem 1. The left PROC implements a 2-way where factor A has 2 levels and factor B has 1 level. Fill in the ? in the right PROC to obtain an equivalent 2-way ANOVA test.

```
proc glm data=study;
   class factorA factorB;
   model y = factorA*factorB;

   contrast 'A1' factorA*factorB  2 2 -1 -1 -1 -1;
   contrast 'A2' factorA*factorB  0 0  1  1 -1 -1;
   contrast 'B'  factorA*factorB  1 -1 1 -1  1 -1;
   contrast 'A1 x B' factorA*factorB  2 -2  -1  1  -1  1;
   contrast 'A2 x B' factorA*factorB  0  0   1 -1  -1  1;
run;
```

```
proc glm data=study;
   class factorB factorA;
   model y = factorB*factorA;

   contrast 'A1' factorB*factorA  ? ? ? ? ? ?;
   contrast 'A2' factorB*factorA  ? ? ? ? ? ?;
   contrast 'B'  factorB*factorA  ? ? ? ? ? ?;
   contrast 'B x A1' factorB*factorA  ? ? ? ? ? ?;
   contrast 'B x A2' factorB*factorA  ? ? ? ? ? ?;
run;
```

---

```
data employees;
 input Dept $ Educ $ Exp Hrs Train Perf Salary;
 ????????
 datalines;
Eng   BS  8 44 35 88 132000
Eng   BS 12 46 40 95 158000
Eng   BS 10 45 30 91 142000
Eng   BS  6 43 28 84 118000
Eng   HS  3 42 20 76  78000
Eng   HS  1 39 15 68  72000
Eng   HS  5 40 25 81  84000
Eng   HS  4 41 22 79  80000
Mktg  BS  8 42 32 87 106000
Mktg  BS 10 43 35 91 116000
Mktg  BS  9 44 38 93 122000
Mktg  BS  6 39 25 80  90000
Mktg  HS  2 37 14 66  61000
Mktg  HS  3 40 16 70  67000
Mktg  HS  4 41 20 75  71000
Mktg  HS  5 38 22 77  74000
Sales BS  9 50 25 90 126000
Sales BS  6 47 20 83 104000
Sales BS 11 52 22 92 134000
Sales BS  7 49 24 86 114000
Sales HS  2 45 10 65  68000
Sales HS  1 42 12 62  62000
Sales HS  4 48 18 79  75000
Sales HS  5 46 15 77  78000
;
run;
```

Expected Output DATA stored as "employees"

| Dept | Educ | Exp | Hrs | Train | Perf | Salary | HS | BS |
|---|---|---|---|---|---|---|---|---|
| Eng | BS | 8 | 44 | 35 | 88 | 132000 | 0 | 1 |
| Eng | BS | 12 | 46 | 40 | 95 | 158000 | 0 | 1 |
| Eng | BS | 10 | 45 | 30 | 91 | 142000 | 0 | 1 |
| Eng | BS | 6 | 43 | 28 | 84 | 118000 | 0 | 1 |
| Eng | HS | 3 | 42 | 20 | 76 | 78000 | 1 | 0 |
| Eng | HS | 1 | 39 | 15 | 68 | 72000 | 1 | 0 |
| Eng | HS | 5 | 40 | 25 | 81 | 84000 | 1 | 0 |
| Eng | HS | 4 | 41 | 22 | 79 | 80000 | 1 | 0 |
| Mktg | BS | 8 | 42 | 32 | 87 | 106000 | 0 | 1 |
| Mktg | BS | 10 | 43 | 35 | 91 | 116000 | 0 | 1 |
| Mktg | BS | 9 | 44 | 38 | 93 | 122000 | 0 | 1 |
| Mktg | BS | 6 | 39 | 25 | 80 | 90000 | 0 | 1 |
| Mktg | HS | 2 | 37 | 14 | 66 | 61000 | 1 | 0 |
| Mktg | HS | 3 | 40 | 16 | 70 | 67000 | 1 | 0 |
| Mktg | HS | 4 | 41 | 20 | 75 | 71000 | 1 | 0 |
| Mktg | HS | 5 | 38 | 22 | 77 | 74000 | 1 | 0 |
| Sales | BS | 9 | 50 | 25 | 90 | 126000 | 0 | 1 |
| Sales | BS | 6 | 47 | 20 | 83 | 104000 | 0 | 1 |
| Sales | BS | 11 | 52 | 22 | 92 | 134000 | 0 | 1 |
| Sales | BS | 7 | 49 | 24 | 86 | 114000 | 0 | 1 |
| Sales | HS | 2 | 45 | 10 | 65 | 68000 | 1 | 0 |
| Sales | HS | 1 | 42 | 12 | 62 | 62000 | 1 | 0 |
| Sales | HS | 4 | 48 | 18 | 79 | 75000 | 1 | 0 |
| Sales | HS | 5 | 46 | 15 | 77 | 78000 | 1 | 0 |

Problem 2.

(A) What SAS code can you write in ???????? to encode HS and BS as dummy variables?

(B) With (A) done, use below to (1) fill in the ??????s and (2) interpret P-value of the HS variable.

```
proc reg data=employees;
    model Salary = HS;
run;
```

| Variable | DF | Parameter_Estimate | Standard_Error | t_Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 111211 | 4484.69502 | 24.80 | <.0001 |
| HS | 1 | -43211 | 9154.34538 | -4.72 | <.0001 |

```
proc reg data=employees;
    model Salary = BS;
run;
```

| Variable | DF | Parameter_Estimate | Standard_Error | t_Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | ?????? | 7980.5732 | 8.52 | <.0001 |
| BS | 1 | ?????? | 9154.34538 | -4.72 | <.0001 |

---

Problem 3.

(A) Write down main effects and interaction contrasts for 2-way ANOVA comparing Dept (factor with 3 levels) vs Educ (factor with 2 levels) on Salary by filling in the PROC GLM below. (Show your work.)

(B) Interpret the p-value of the interaction contrast "Eng vs Mktg (BS vs HS)" shown.

(C) Compute by hand the missing Mean_Square and F_Value. (Leave answers as +, x, ÷ of numbers.)

```
proc glm data=employees;
    class Dept Educ;
    model Salary = Dept Educ Dept*Educ;
    contrast 'Dept: Eng vs Mktg'    Dept
    contrast 'Dept: Mktg vs Sales'  Dept
    contrast 'Educ: BS vs HS'       Educ
    contrast 'Eng vs Mktg (BS vs HS)'   Dept*Educ
    contrast 'Mktg vs Sales (BS vs HS)' Dept*Educ
run;
```

| Contrast | DF | Contrast_SS | Mean_Square | F_Value | Pr > F |
|---|---|---|---|---|---|
| Dept: Eng vs Mktg | 1 | 1540562500 | 1540562500 | 12.13 | 0.0027 |
| Dept: Mktg vs Sales | 1 | 182250000 | | | 0.2464 |
| Educ: BS vs HS | 1 | 14602666667 | 14602666667 | 115.01 | <.0001 |
| Eng vs Mktg (BS vs HS) | 1 | 351562500 | 351562500 | 2.77 | 0.1134 |
| Mktg vs Sales (BS vs HS) | 1 | 72250000 | 72250000 | 0.57 | 0.4604 |

Also study:
- PROC GLM, REG, MEANS,
- Diagnostic plots for above PROCs
- How to fill in ANOVA table by hand

Ignore:
- Lessons 9, 11, 15,
- PROC: FORMAT, FREQ

**Problem 4.**

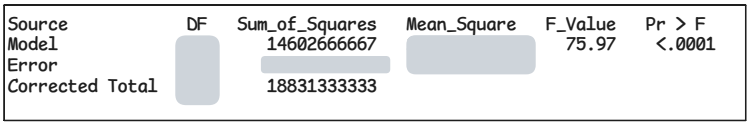
(A) Find DF

(B) Find SSE

(C) Interpret P-value

(D) Rewrite this PROC as PROC REG

```
proc glm data=employees;
        class Educ;
    model Salary = Educ;
run;
```



| Source | DF | Sum_of_Squares | Mean_Square | F_Value | Pr > F |
|---|---|---|---|---|---|
| Model | | 14602666667 | | 75.97 | <.0001 |
| Error | | | | | |
| Corrected Total | | 18831333333 | | | |